Large but Rank-Deficient Analysis Operator Learning

Mehrdad Yaghoobi, Mike E. Davies
Institute for Digital Communications (IDCOM), The University of Edinburgh
AGB, King's Buildings, Mayfield Road, Edinburgh, EH9 3JL
Emails: {m.yaghoobi-vaighan,Mike.Davies}@ed.ac.uk

***Abstract*—The problem of analysis operator learning (AOL) for medium to large-scale signal blocks will be investigated here. Some signal structures can only be discovered/learned if we consider a large block-size for the signals. The goal here is to show that learning an operator does not necessarily need to to be done in such a high dimensional space. We present a new technique based on the projection onto the most significant K-dimensional singular space of the training samples, $K \ll n$, to reduce the computational complexity of the learning algorithm[1].**

I. Introduction

The AOL problem has recently received more attention, after introduction of a new low-dimensional signal model, called the *cosparse* model [1]. The (linear) analysis operator maps the signals to the analysis space, where q elements are (approximately) zero, *i.e.* (approximately) q-cosparse. In a finite dimensional setting, such a linear operator can be represented with a tall matrix, which its rows are called the analysers. Accordingly, we *can* formulate AOL with an optimisation problem to find an operator $\mathbf{\Omega} \in \mathbb{R}^{a \times n}$, which cosparsifies a given set of exemplars $\mathbf{Y} \in \mathbb{R}^{n \times L}$ [2] as follows,

$$\min_{\mathbf{\Omega}} \|\mathbf{\Omega Y}\|_1, \text{ s.t. } \mathbf{\Omega} \in \mathcal{C} \tag{1}$$

where $\mathcal{C}$ is a constraint on the singular values of $\mathbf{\Omega}$ and its row norms, *i.e.* $\mathcal{C} = \{\mathbf{\Omega} \in \mathbb{R}^{a \times n} \ : \ \mathbf{\Omega}^T\mathbf{\Omega} = \mathbf{I}, \ \forall i \ \|\omega_i\|_2 = \sqrt{\frac{a}{n}}\}$. As this problem has a non-smooth but convex objective, which is supposed to be minimised over the intersection of two matrix manifolds, we find a local optimum of (1) using a sub-gradient descent type algorithm. [3] introduces a relaxation for the constraint to accelerate the convergence of algorithm. However, computational complexity of the relaxed formulation is still high for large-scale problems. As an example, if we learn an operator for 8×8 image patches with 8^4 training samples, Relaxed AOL [3] only takes 30 seconds on a 2.6 GHz, 12 core Intel processor workstation using Matlab environment, while such a learning for 32×32 patches and $2 {\times} 32^3$ training samples takes more than two and a half days!

A. Low-dimensional Analysis Operator Learning

If there exists a low-dimensional structure in the training matrix $\mathbf{Y}$, we can reduce the computational complexity of the algorithm by learning in such a low-dimensional space and map the operator back to the original space. Let $\mathbf{Y} \triangleq \mathbf{U}_{n\times n}\Sigma_{n \times n}(\mathbf{V}_{L\times n})^T$ be a Singular Value Decomposition (SVD) of $\mathbf{Y}$. The most significant K-dimensional singular space can be found by keeping the largest K singular values and setting the rest to be zero. If Σ_K is such a diagonal matrix, $\mathbf{Y}_K = \mathbf{U}\Sigma_K\mathbf{V}^T$ is the closest rank K, $n \times L$ matrix to the training corpus $\mathbf{Y}$. When $\mathbf{Y}_K$ is a good approximation for $\mathbf{Y}$, which is what we observed for the natural image training samples and a reasonable selection of K, we can reduce the dimensionality of learning problem (1) by using $\widetilde{\mathbf{Y}}_{K\times L} \triangleq \{\mathbf{U}^T\mathbf{Y}\}^{(K)}$, where $\{\cdot\}^{(K)}$ is a shrinking operator which selects the first K rows of the operand, as the training matrix and find the solution $\widetilde{\mathbf{\Omega}}^*$. We should now map $\widetilde{\mathbf{\Omega}}^*$ to the original space using $\mathbf{\Omega}^* = \widetilde{\mathbf{\Omega}}^*\mathbf{U}^T$, to find a solution $\mathbf{\Omega}^*$ for (1).

[1]This work was supported by EU FP7, FET-Open grant number 225913 and EPSRC grant EP/J015180/1.

II. Simulation Results

We chose some canonical images, *i.e. Barbara, boat, Lena, fingerprint, flinstones, house* and *peppers*, to learn the operator. The image patches were randomly located in the reference images with 32×32 ($n = 1024$), $a = 2n$, $K = 16^2$ and $L = 16^4$ and the RelaxedAOL algorithm [3] iterated 10^5 times. We have shown the first K singular vectors of $\mathbf{Y}_{:,1:8000}$, as some 32×32 images in Figure 1, while they capture more than 96% of the signal energy. When we project $\mathbf{Y}$ onto this K-dimensional space and learn the operator, we will find a rank-K operator after 22 hours using the mentioned machine, which is shown in Figure 2, *i.e.* each 1024 length analyser reshaped to a 32×32 image. We observe that the learned analysers have some spatial-locality and a 2D Gabor structure. This is indeed an interesting fact as we do not see such a structure in the (left) singular vectors. Such a structure can not be explored in the small-size problems, *e.g.* canonical 8×8 patch size.

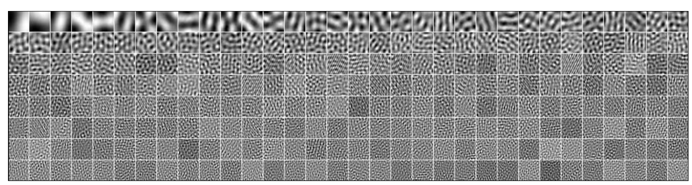

Fig. 1. The first 256 singular vectors of the training matrix.

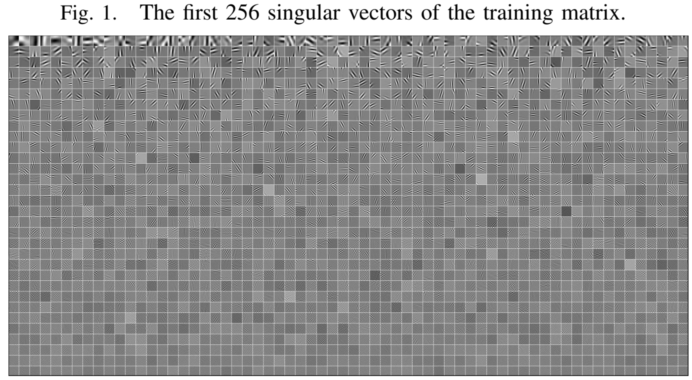

Fig. 2. The learned analysers for the selected 256-dimensional singular space of Figure 1.

References

[1] S. Nam, M. E. Davies, M. Elad, and R. Gribonval, "The cosparse analysis model and algorithms," *Applied and Computational Harmonic Analysis*, vol. 34, no. 1, pp. 30–56, 2013.

[2] M. Yaghoobi, S. Nam, R. Gribonval, and Davies M., "Constrained overcomplete analysis operator learning for cosparse signal modelling," *IEEE Trans. on Signal Processing*, vol. 61, no. 9, pp. 2341–2355, 2013.

[3] M. Yaghoobi and M.E. Davies, "Relaxed analysis operator learning," in *NIPS, Workshop on Analysis Operator Learning vs. Dictionary Learning: Fraternal Twins in Sparse Modeling*, 2012.